Operations Research Perspectives 1 (2014) 6–17

Contents lists available at ScienceDirect

Operations Research Perspectives

journal homepage: www.elsevier.com/locate/orp

Using the latent class approach to cluster firms in benchmarking: An application to the US electricity transmission industry

Manuel Llorca [a,*], Luis Orea [a], Michael G. Pollitt [b]

[a] Oviedo Efficiency Group, Department of Economics, University of Oviedo, Spain

[b] Energy Policy Research Group and Judge Business School, University of Cambridge, United Kingdom

ARTICLE INFO

Article history:
Received 29 January 2014
Received in revised form
26 February 2014
Accepted 18 March 2014
Available online 13 April 2014

JEL classification:
D22
L51
L94

Keywords:
Electricity transmission
Utilities regulation
Latent class model approach
Nonparametric analysis

ABSTRACT

In this paper we advocate using the latent class model (LCM) approach to control for technological differences in traditional efficiency analysis of regulated electricity networks. Our proposal relies on the fact that latent class models are designed to cluster firms by uncovering differences in technology parameters. Moreover, it can be viewed as a supervised method for clustering data that takes into account the same (production or cost) relationship that is analysed later, often using nonparametric frontier techniques. The simulation exercises show that the proposed approach outperforms other sample selection procedures. The proposed methodology is illustrated with an application to a sample of US electricity transmission firms for the period 2001–2009.

1. Introduction

Electricity networks are often regulated by implementing incentive-based regulation schemes that use some types of benchmarking, i.e. a comparison of utilities' performance with best-practice references. As shown by Zhou et al. [1], the nonparametric DEA (Data Envelopment Analysis) has become a very popular tool in energy and environmental studies, especially for benchmarking electric utilities. Unlike the econometric SFA (Stochastic Frontier Analysis) that requires the specification of a particular functional form for the cost or production functions to be estimated, DEA imposes fewer assumptions on the shape of firms' technology and it allows regulators to address traditional convergence problems and the well-known 'wrong skewness problem' in the SFA literature.

A key issue that is sometimes not taken into account by regulators (and researchers) is the heterogeneity or unobserved differences among firms, although utilities are usually quick to mention this issue to the regulators. This concern underlies the negotiations between regulators and utilities, where utilities wield uniqueness as a reason to avoid being compared with their peers. However, it is often assumed in this setting that the whole set of benchmarked firms share the same technology, and hence differences in behaviour are attributed to inefficient use of factors that are under the control of the companies. Possible differences among utilities associated with different technologies are either overlooked or are addressed using simple sample selection procedures, mostly based on factors that may affect performance such as geographic location or utilities' size. Therefore, the efficiency scores obtained from these analyses might be biased and some firms might be penalized (or rewarded) in excess if their underlying technology is less (more) productive than the technology used by other firms operating with more (less) advantageous conditions. This is particularly important in the case of incentive regulation and benchmarking of electricity networks where the results of efficiency analysis have important financial implications for the firms.

In this paper we examine whether we should (a) split the sample arbitrarily on the basis of a single size variable, or (b) use a

* Correspondence to: Faculty of Economics and Business, University of Oviedo, Av. del Cristo, s/n, 33006, Oviedo, Spain. Tel.: +34 985104885.
E-mail addresses: llorcamanuel@gmail.com, llorcamanuel@uniovi.es (M. Llorca).

http://dx.doi.org/10.1016/j.orp.2014.03.002

comprehensive statistical procedure to control for technological differences, before carrying out a traditional efficiency analysis of regulated electricity networks. We advocate using the so-called latent class model (LCM) approach that allows us to split the electricity networks into a number of different classes, where each class is associated with a different technology. We advocate this approach for several reasons. First, LCM clusters firms by searching for differences in production or cost parameters, which is exactly what regulators are looking for. Second, our approach can be viewed as a "supervised" method for clustering data as it takes into account in the first stage the same (production or cost) relationship that is analysed later, often using nonparametric frontier techniques. Indeed, the literature on data dimension reduction uses this expression for those methods that not only use the information contained in the explanatory variables to be aggregated, but also the information of the dependent variable that will be predicted later on. And third, our approach is not more "technical" than other clustering methods as it can be implemented using standard software and using the same variables that will be used to get efficiency scores in a later stage. Having practicality in mind, we have proposed some simplifications such as the use of simple specifications for both the deterministic (e.g. Cobb–Douglas) and stochastic (e.g. normal distribution) parts of the model to facilitate its application. The use of the same variables in both the latent class stage and the second, DEA, stage also contributes to simplify the use of the proposed procedure.

The same idea is currently being developed by Agrell et al. [2] in a very recent study where they use the LCM approach to control for technological differences in an application to Norwegian power distribution firms. Our paper reinforces the approach from both a theoretical and an empirical point of view. In particular, we carry out a simulation analysis to examine whether the latent class approach outperforms other more arbitrary and less robust procedures for splitting a sample of observations—such as the k-means clustering algorithm or simply using the median of some relevant variables. The simulation exercises confirm our expectations and show that the proposed approach outperforms alternative sample selection procedures. We illustrate this procedure with an application to the US electricity transmission firms examined in [3]. We find two statistically different groups of firms that should be compared or treated separately. In order to confirm the results from the simulation exercise, we compare the partition of the sample obtained through this method with those from alternative clustering procedures.

This paper is organized as follows. Section 2 introduces the two-stage procedure that is proposed to control for unobservable differences in firms' technology (environment) in energy regulation. Section 3 introduces the simulation analysis performed and its main outcomes. Section 4 uses data from the US electricity transmission industry to compare the relative performance of our approach and alternative procedures. Section 5 concludes.

2. A two-stage procedure to address unobserved heterogeneity in utility regulation

As Haney and Pollitt [4] pointed out in a recent survey, regulators have been using several statistical methods to determine the performance of energy utilities. Obtaining reliable measures of firms' performance requires dealing with controllable factors and monitoring for the different environmental conditions under which firms operate. However, both regulators' reports and academic studies do not usually deal with these technological differences. Statistical methods have recently been developed to address this issue. In most of these methods, heterogeneity is understood as an unobserved determinant of the production/cost frontier, while inefficiency is interpreted as the 'distance' to the frontier once heterogeneity has been taken into account.

Following Greene [5,6] we can distinguish two types of models that allow us to achieve our aim, namely the so-called True Fixed Effects (TFE) and True Random Effects (TRE) models introduced by this author, and the LCM, also known as finite mixture models, which have been broadly used in several fields of research (see [7]; or [8], for simple applications; and [9]; or [10], for more comprehensive applications that aim to examine technological gaps using a metafrontier approach). Both approaches have their own strengths and weaknesses. In the TFE/TRE models, unobserved heterogeneity is captured through a set of firm-specific intercepts that are simultaneously estimated with other parameters. Hence, this approach assumes that there are as many technologies as firms. However, as it imposes common slopes for all firms, all of them share the same marginal costs, economies of scale and other technological characteristics.

In contrast to the TFE/TRE models, the LCM approach allows the estimation of different parameters for firms belonging to different groups. This can be easily seen if the general specification of a cost function in this framework is expressed as follows:

$$\ln X_{it} = \alpha_j + \beta_j \ln Y_{it} + v_{it|j} \tag{1}$$

where i stands for firms, t for time and $j = 1, \ldots, J$ for class. X_{it} is a measure of firms' cost, Y_{it} is a vector of explanatory variables, and the random term v_{it} follows a normal distribution with zero mean and variance σ_v^2. As both α_j and β_j, are j-specific parameters, the technological characteristics vary across classes.

Letting θ_j denote all parameters associated with class j, the conditional likelihood function of a firm i belonging to class j is $LF_{ij}(\theta_j)$. The unconditional likelihood for firm i is then obtained as the weighted sum of their j-class likelihood functions, where the weights are the probabilities of class membership, P_{ij}. That is:

$$LF_i(\theta, \delta) = \sum_{j=1}^{J} LF_{ij}(\theta_j) P_{ij}(\delta_j), \quad 0 \le P_{ij}(\delta_j) \le 1,$$

$$\sum_{j=1}^{J} P_{ij}(\delta_j) = 1 \tag{2}$$

where $\theta = (\theta_1, \ldots, \theta_j), \delta = (\delta_1, \ldots, \delta_j)$ and the class probabilities are parameterized as a multinomial logit model:

$$P_{ij}(\delta_j) = \frac{\exp(\delta_j' q_i)}{\sum_{j=1}^{J} \exp(\delta_j' q_i)}, \quad j = 1, \ldots, J, \qquad \delta_J = 0 \tag{3}$$

where q_i is either an intercept or a vector of individual-specific variables. Therefore, the overall likelihood function resulting from (2) and (3) is a continuous function of the vectors of parameters θ and δ, and can be written as:

$$\ln LF\,(\theta, \delta) = \sum_{i=1}^{N} \ln LF_i\,(\theta, \delta) = \sum_{i=1}^{N} \ln \left\{ \sum_{j=1}^{J} LF_{ij}\left(\theta_j\right) P_{ij}\left(\delta_j\right) \right\}. \tag{4}$$

Maximizing the above maximum likelihood gives asymptotically efficient estimates of all parameters. A necessary condition to identify the whole set of parameters is that the sample must be generated from at least two different technologies or two noise terms.

Several comments are in order. First, in this framework each firm belongs to one and only one class.[1] Therefore, the probabilities

[1] This does not mean that a specific firm is going to be always in the same class. The clusters are created without taking into account the panel structure of the data,

of class membership just reflect the uncertainty that researchers or regulators have about the true partition of the sample. The estimated parameters can be used to compute posterior class membership probabilities using the following expression:

$$P(j|i) = \frac{LF_{ij}\left(\hat{\theta}_j\right) P_{ij}\left(\hat{\delta}_j\right)}{\sum_{j=1}^{J} LF_{ij}\left(\hat{\theta}_j\right) P_{ij}\left(\hat{\delta}_j\right)}. \quad (5)$$

These posterior probabilities of membership can then be used to allocate each firm to a particular class, e.g., each firm is allocated to the class with the higher posterior probability.

Second, only between-groups and not individual heterogeneity is controlled using a latent class model because all firms belonging to a particular group share the same technology. This situation is possible in energy economics if firms operating in areas with different environmental conditions must choose between a limited number of technical standards[2] to expand and maintain their networks. If firms have similar technologies, the estimated differences in technology (i.e. parameters) are likely to be capturing heterogeneity in operating environments.[3] Therefore, the differences in parameters between classes can be interpreted either as differences in technology or differences in environmental variables that might be unobserved.

Third, the number of classes J should be chosen in advance by the researcher or regulator. Selecting the number of classes is a key issue of the proposed approach, and is common to other clustering methods. Fortunately there are several statistical tests that are commonly used and accepted in the finite mixture models literature to choose the appropriate number of classes. For instance, the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) are frequently used in the LCM literature. These criteria involve minimizing an index that balances the lack of fit (too few classes) and overfitting (too many classes) as it includes a penalty that increases with the number of parameters. Models with lower AIC or BIC are generally preferred. The BIC considers a greater penalty for overfitting than AIC and, hence, BIC tends to favour more parsimonious models, which in turn help to estimate model coefficients with more precision (see [11], p. 61). Many authors (see for instance Koehler and Murphree [12]) observed that the traditional AIC criterion and some of their variants tend to overestimate the correct number of classes. For those criteria that tend to overfit and favour more comprehensive models, it is very useful to examine a graph of the values of the computed statistic as the classes increase and look for the natural bend or break point where the curve flattens out. The number of data points till the "break" (i.e., including the point at which the break occurs) can be used as the number of classes to select. This method (labelled a "scree test") is often used in principal components or factor analyses to select the number of factors and it is described and pictured in every textbook discussion of factor analysis (see, for instance, [13], p. 2–3).

Finally, it should be noted that the random term in (1) follows a symmetric distribution because it does not include a traditional one-sided inefficiency term. In other words, we advocate using a simple normal distribution in the first stage of our procedure and obtaining the efficiency scores later. There are three reasons for this. First, ignoring the asymmetric error term traditionally associated with inefficiency prevents the appearance of convergence problems in practice when estimating a latent class model, which by nature is highly non-linear. This facilitates replication of the procedure when researchers or regulators compare different specifications of the underlying technology. Second, this empirical strategy allows us to compute efficiency scores using more flexible representation of firms' technologies if nonparametric techniques such as DEA are employed. Finally, DEA is the method mainly used by regulators (see [4]).

The main advantage of using an LCM approach to cluster firms is that it allows us to control for environmental factors (i.e. contextual z-variables) that are not observable, difficult to measure accurately or even unknown in some cases. The LCM–DEA approach also allows the inclusion of z-variables to identify groups of comparable firms that share similar environmental or technological features (for a discussion on this topic in the DEA and SFA literature, see for instance [14,15]). Thus it is more sophisticated than simply including z-variables without clustering. In this sense, our approach is consistent with the idea of benchmarking, which is based on the existence of comparable firms. However it extends this by avoiding the need for arbitrary clustering, which is often undertaken by researchers and regulators. Under arbitrary clustering larger samples are often split into sub-samples to be analysed separately on the basis of a single size metric (such as number of customers) or using subjective value judgement.

In a second stage DEA is separately applied for each class. DEA is a type of efficiency analysis which involves mathematical programming to construct a frontier of best performing companies.[4] Farrell [16] was the first to propose this type of frontier analysis and since then there have been many authors who have developed and applied different models which have enlarged the literature in DEA methodology (see [17]).

In this paper, we will use an input-oriented DEA model as we assume that the output level cannot be modified by firms. This is a reasonable assumption for a network utility required to provide network capacity to service ultimate demand which is largely out of its control. Technical inefficiency can be then viewed as a proportional reduction in input usage or cost while maintaining the output levels constant. In our simulation exercise we impose constant returns to scale (CRS) as similar results are obtained if this assumption is relaxed. The optimization problem in this case can be represented as:

$$\begin{aligned} &\min_{\theta,\lambda} \theta, \\ &st \; -y_i + Y\lambda \geq 0, \\ &\quad \theta x_i - X\lambda \geq 0, \\ &\quad \lambda \geq 0 \end{aligned} \quad (6)$$

where λ is a vector of constants and θ is a scalar calculated for each observation which represents the efficiency score for the ith

i.e. a particular firm can be in different clusters over time. It has been done in this way to give more flexibility to our model by allowing changes in firms' technology along the sample period. Moreover this type of approach usually yields similar results to those obtained if the belonging to a certain class is imposed for the whole sample period.

2 These standards are either proposed by the International Electrotechnical Commission or the Institute of Electrical and Electronics Engineers.

3 We are grateful to a referee for pointing this out.

4 Although DEA is a rather flexible method that does not impose implicitly the same 'parameters' on the whole sample of firms, we would like to point out, however, that obtaining different marginal products (elasticities) at different points of the sample does not mean in economics that we have estimated different technologies. From the engineering point of view, the term "technology" is often associated with a particular production process. Nevertheless in microeconomic theory, technology is more broadly defined as the set of processes technically feasible and available for firms in a moment of time, and movements along the frontier just represent different production processes within a certain technology. The differences in technology across firms (or over time) are captured in economic analysis by shifts of the production frontier. Thus, only differences in production processes are controlled when a single frontier is estimated either using DEA or SFA.

firm. y_i and x_i are the vectors of outputs and inputs for the ith firm respectively, while Y and X are the output and input matrices for all I firms. This linear programming problem must be solved I times and gives an efficiency score θ equal or lower than one for each firm. In our empirical application we do not assume that all the firms exhibit constant returns to scale as electricity transmission firms are natural monopolies and increasing returns to scale were obtained in many applied studies.[5] A variable returns to scale (VRS) specification only requires adding the convexity constraint $I1'\lambda = 1$ to the minimization problem in (6). $I1$ is a vector of ones, and multiplying by the vector of weights λ ensures that firms are only compared with firms of a similar size.

As pointed out by one referee, we are using a two-stage procedure that combines parametric and nonparametric techniques. Although it is unlikely that both techniques are fully compatible, we do not have to deal with the inconsistency problem that appears in the traditional two-stage DEA procedure. This problem arises in a different situation, when DEA is undertaken first and parametric analysis is then performed on the DEA results. Simar and Wilson [21] have shown that applying a parametric regression in a second stage using the estimates obtained in a first stage through DEA is not consistent because firms' inefficiency is a relative measure and, hence, the nonparametric efficiency scores are serially correlated. As the order is reversed, this problem does not emerge in our case. Moreover, none of the variables used in our second stage are predicted or estimated variables.

Furthermore it should be mentioned that the DEA approach can be used as a clustering method. There is an evolving literature on this topic from [22–24]. However [25] find that this approach, which is based on the piecewise production functions obtained from DEA models for clustering the data, faces several problems. Firstly, it may have alternative optimal solutions and hence the clusters produced are not unique. Secondly, they find that it is possible not to achieve any strictly positive multiplier weight for inputs and outputs in evaluating all firms. And finally, some of the obtained clusters may have overlapping units.

3. Simulation analysis

In this section we carry out a simulation exercise to examine whether a latent class approach is a good procedure to find groups of comparable companies within a sample when we aim to apply a benchmarking with DEA, commonly used in regulatory processes. It should be pointed out that the LCM is compared with other clustering methods as a point of comparison. However, the main objective in our simulation is to test the discriminatory power of the LCM under different scenarios when technological and output differences arise, which as far as we know has not been performed before in the efficiency analysis literature.

The simulation exercise can be summarized as follows. Firms' costs are calculated using simulated data and following the normalized linear specification proposed by Bogetoft and Otto [26] for the regulation of electrical Distribution System Operators in Germany. This functional form allows us to easily introduce heteroscedasticity in our data generation process. Following this specification, our cost function can be expressed as follows:

$$\frac{X_i}{Y_{1i}} = \beta_1 + \beta_2 \frac{Y_{2i}}{Y_{1i}} + u_i^+ + v_i \tag{7}$$

where X_i is our cost, while β_1 and β_2 stand for the marginal costs of the outputs Y_1 and Y_2 and define our technologies. Although we are imposing constant returns to scale in (7) to prevent size effects when comparing our sample separating methods, the use of variable returns to scale in the simulation produces the same partition of the sample and slightly larger efficiency scores.

In the papers in which simulations are carried out, the choice of the approach used in the Data Generation Process (DGP) is frequently quite contentious (see for instance [27]). However, the way in which our DGP is defined here is not uncommon in efficiency analysis papers and can be found both in the SFA literature (see for instance [28]; or [29]) and in the DEA literature (see for instance [30]; or [31]).

Inefficiency levels are obtained assuming that the inefficiency term, u^+, is a positive half-normal distribution with zero mean and σ_u^2 variance. Random noise is simulated assuming that the noise term v follows a normal distribution with zero mean and σ_v^2 variance. We impose $\sigma = \sqrt{\sigma_u^2 + \sigma_v^2}$ equal to 1, which, given the specification that we have chosen, implies that the size of the random term in our function is relatively low, i.e. our levels of generated efficiency are quite high. We also fixed $\gamma = \sigma_u^2/(\sigma_u^2 + \sigma_v^2)$ equal to 0.5, which implies that the weights of inefficiency and noise in the function are the same. Given the previous values, this implies that $\sigma_u = \sigma_v = 0.71$, and therefore is equivalent to generating a value of $\lambda = \sigma_u/\sigma_v$ equal to 1.[6]

We randomly generate 1000 observations of two hypothetical outputs (Y_1, Y_2) using a uniform distribution between 0 and 1. We have chosen this distribution instead of the normal distribution because these variables cannot take negative values, and outputs in DEA must be positive. As the random noise term takes both positive and negative values, we impose on all technologies that $(\beta_1 + \beta_2) = 10$ to obtain positive costs. Technologies thus differ in relative marginal costs, i.e. the relative weight of each β. In particular, we have simulated three possible technologies:

- Technology A: $\beta_2 = \beta_1$, $(\beta_1 = 5, \beta_2 = 5)$
- Technology B: $\beta_2 = 2\beta_1$, $(\beta_1 = 10/3, \beta_2 = 20/3)$
- Technology C: $\beta_2 = 4\beta_1$, $(\beta_1 = 2, \beta_2 = 8)$.

Both coefficients are the same in technology A, while marginal costs are increasingly different in the other two technologies, B and C. Although these differences in parameters between classes are associated with different technologies, we have already mentioned before that they can be interpreted either as differences in technology or differences in environmental variables. Next, we will examine the robustness of our results by adding differences between outputs. In particular, we modify the original statistical distribution of the second output by doubling $(Y_2 \sim 2 \cdot U(0, 1))$ and quadrupling $(Y_2 \sim 4 \cdot U(0, 1))$ its range of values.

Taking into account that we always apply the technology A to the first 500 observations and then B or C to the following 500 observations, and that we have three output distributions, 6 possible scenarios are obtained. In Table 1, we show the scenarios and the percentage success in predicting the underlying class membership using different clustering methods. Percentages of success can be obtained through the identification of the groups, which is possible after comparing the real β-ratios with those obtained using group-specific OLS regressions. The estimated ratios that are also shown in Table 1 give an idea about how well each procedure is able to identify the underlying, but different, technologies.

The first empirical exercise has to do with the case in which DEA is applied using the real separation of our data. By construction, the percentage of success in this case is 100%. For this reason, this exercise is used as a benchmark to study the performance of four

[5] See for instance [18–20,3].

[6] Although the values of these parameters have been arbitrarily chosen, the results obtained from the simulation are consistent with respect to changes in them as long as we keep the underlying efficiency at 'normal' levels.

Table 1
First stage simulation results: percentage success in identifying technologies.

Simulation	Procedure	% Success	Underlying technology	
			Group 1 ($\beta 1/\beta 2$)	Group 2 ($\beta 1/\beta 2$)
A&B	Simulation	–	1.000	0.500
(OD 1)	Real separation	100.00	1.080	0.597
	Median (C)	49.60	0.890	0.756
	Cluster (Y_1, Y_2)	46.50	0.849	0.815
	Cluster (Y_1, Y_2, C)	49.30	0.894	0.763
	LCM	65.70	1.063	0.555
A&C	Simulation	–	1.000	0.250
(OD 1)	Real separation	100.00	1.080	0.331
	Median (C)	50.20	0.678	0.575
	Cluster (Y_1, Y_2)	46.50	0.646	0.642
	Cluster (Y_1, Y_2, C)	49.90	0.684	0.593
	LCM	79.20	1.162	0.337
A&B	Simulation	–	1.000	0.500
(OD 2)	Real separation	100.00	1.077	0.597
	Median (C)	55.00	0.834	0.799
	Cluster (Y_1, Y_2)	54.00	0.822	0.812
	Cluster (Y_1, Y_2, C)	55.10	0.822	0.802
	LCM	79.30	1.110	0.596
A&C	Simulation	–	1.000	0.250
(OD 2)	Real separation	100.00	1.077	0.331
	Median (C)	57.20	0.723	0.562
	Cluster (Y_1, Y_2)	54.00	0.656	0.609
	Cluster (Y_1, Y_2, C)	58.30	0.714	0.529
	LCM	87.90	1.099	0.337
A&B	Simulation	–	1.000	0.500
(OD 3)	Real separation	100.00	1.076	0.598
	Median (C)	57.40	0.868	0.765
	Cluster (Y_1, Y_2)	53.90	0.833	0.785
	Cluster (Y_1, Y_2, C)	57.80	0.863	0.754
	LCM	90.60	1.097	0.583
A&C	Simulation	–	1.000	0.250
(OD 3)	Real separation	100.00	1.076	0.331
	Median (C)	60.60	0.779	0.493
	Cluster (Y_1, Y_2)	53.90	0.674	0.576
	Cluster (Y_1, Y_2, C)	61.80	0.772	0.486
	LCM	94.70	1.102	0.328

sample separation methods: the median of the cost, the k-means clustering algorithm considering the outputs, the k-means clustering algorithm including both outputs and cost, and the latent class model (that involves both output and cost information). Looking at the percentages of success and the β-ratios we can confirm that the LCM is the method that better allocates observations to specific technologies. It is also the best clustering method at identifying the relationship between technologies represented by the β-ratios. As we move to a different scenario where there are more uneven features among groups, we observe that there is a clear divergence in the behaviour of the procedures: whereas the LCM improves its percentage of prediction success,[7] the alternative procedures only slightly improve their performances.

We show in Table 2 the average efficiencies that are obtained after DEA is applied separately to each group of firms. The last column shows the sum of squared differences (SSD) with respect to the real separation case. The SSD is calculated as the total sum of squared differences of the predicted efficiency with respect to the value of the underlying efficiency of each observation. The smallest SSD value allows us to identify the best individual predictor procedure, i.e. the clustering method that better predicts the 'real' efficiencies. Leaving aside the real separation case where SSD is zero by construction, LCM provides by far the smallest SSD in all scenarios. As LCM is the procedure that gives the closest efficiency levels to the real separation case, it is the best at predicting individual efficiencies.

When we move from a model with only one class to a model with two classes and unobserved heterogeneity is somehow taken into account (or 'removed') in a first stage, larger efficiency scores are obtained when carrying out a traditional DEA analysis. The computed efficiency improvements are partially caused by the fact that the number of peers necessarily decreases when a model with two classes is used,[8] regardless of the clustering method. However, our simulation exercise shows that these efficiency increases have also to do with the selection of a specific clustering method. In particular, Tables 1 and 2 (and Fig. 1) indicate that the better the partition is, the larger the average efficiency scores are. Moreover, this result happens regardless of whether we carry out either traditional DEA or SFA (not shown) analyses in the second stage of our procedure.

Fig. 1 shows the positive correlation that exists between efficiency and success in assigning observations to technologies using the LCM approach. This figure allows us to examine the discriminatory power of the model when there are either larger differences between technologies (illustrated as the shift from the blue to the red line) or between output data generation

[7] The estimated probabilities for the most likely latent class also increase, so the LCM not only improves its prediction capacity but also the precision with which each observation is assigned.

[8] This does not necessarily happen when we move from 2 to 3 classes (and so on) because there is a reallocation of the observations into the different classes. Indeed, as a larger partition of the sample does not imply that one class is divided into two separable classes, some observations might have "new" peers and, hence, their (relative) efficiency might be less than before.

Table 2
Second stage (DEA) simulation results: predicted vs. underlying efficiency.

Simulation	Procedure	Av. Eff. (%)	Sum of squared differences
A&B	Real separation	76.73	–
(OD 1)	No separation	67.15	135,252
	Median (C)	73.29	118,885
	Cluster (Y_1, Y_2)	70.38	108,590
	Cluster (Y_1, Y_2, C)	72.91	118,993
	LCM	73.35	40,268
A&C	Real separation	75.16	–
(OD 1)	No separation	54.70	533,029
	Median (C)	65.26	322,549
	Cluster (Y_1, Y_2)	61.65	379,683
	Cluster (Y_1, Y_2, C)	64.65	338,483
	LCM	78.93	139,993
A&B	Real separation	83.61	–
(OD 2)	No separation	73.28	151,038
	Median (C)	76.75	121,264
	Cluster (Y_1, Y_2)	75.93	119,387
	Cluster (Y_1, Y_2, C)	76.22	124,400
	LCM	85.75	51,232
A&C	Real separation	83.14	–
(OD 2)	No separation	63.22	507,703
	Median (C)	69.29	372,090
	Cluster (Y_1, Y_2)	68.14	386,773
	Cluster (Y_1, Y_2, C)	68.32	389,718
	LCM	85.87	45,663
A&B	Real separation	89.26	–
(OD 3)	No separation	78.75	180,309
	Median (C)	80.36	158,268
	Cluster (Y_1, Y_2)	79.99	160,646
	Cluster (Y_1, Y_2, C)	80.20	159,799
	LCM	90.76	29,598
A&C	Real separation	89.24	–
(OD 3)	No separation	70.49	511,481
	Median (C)	73.45	429,212
	Cluster (Y_1, Y_2)	72.86	446,596
	Cluster (Y_1, Y_2, C)	72.95	438,210
	LCM	90.15	16,511

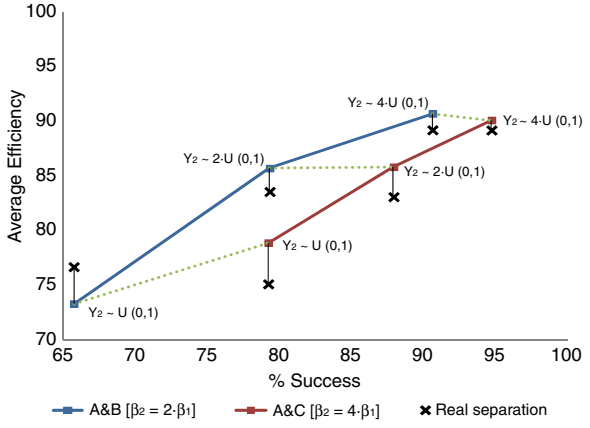

Fig. 1. Average efficiency and percentage of success for the LCM.

processes (illustrated as movements along the red and blue lines). As expected, the percentages of success are much larger when the two technologies differ notably in their characteristics. It is worth mentioning that this increase in percentages of success is especially important when there is no separating information on the output side, i.e. when both outputs are similarly generated. When outputs provide additional information to split the sample, both efficiency levels and percentages of success increase, regardless of whether the technologies are similar or diverse. On the other hand, Fig. 1 also shows that as inequalities between groups rise, the average efficiency score obtained using LCM as a sample separation method even exceed the average efficiency score from the real separation case. This shows that an imperfect assignment of firms to groups can lead us to obtain higher levels of efficiency. In other words, making a good partition of the sample does not necessarily imply obtaining larger efficiencies.

In summary, the above results clearly indicate that LCM deals with unobserved heterogeneity much better than the other clustering methods. We attribute this better performance to the fact that LCM splits the data taking into account the objective of the second stage, where a relationship between outputs and inputs (or costs) is estimated in order to compute inefficiency scores. In this sense, and borrowing the terminology used for dimension reduction, this approach can be interpreted as a 'supervised' method to split the data.

From a regulation point of view, the above results suggest that, given a number of classes, regulators could use this statistic (i.e. the mean efficiency) to compare the relative performance of several clustering methods in a real case in which they do not have information about the 'underlying partition' of the sample. Our proposed procedure thus can be labelled as a conservative approach. However, using a method that provides conservative efficiency estimates is common among regulators. For example, in Germany, the regulator assesses the performance of each firm using both DEA and SFA efficiency scores and chooses the larger of the two estimates [32]. Here we provide an additional reason, based on simulation results, that justifies the use of a conservative approach when, and only when, clustering methods are used in benchmarking.

4. Application to the US electricity transmission industry

4.1. Data, sample and variables

We next illustrate the proposed procedure with an application to the US electricity transmission industry. As is highlighted by

Table 3
Descriptive statistics.

	Variable	Units	Mean	Max.	Min.	Std. Dev.
Totex	Cost	US$	144,602,000	667,127,000	20,713,600	120,324,000
Peak load	Output	MW	6173	23,111	380	5533
Electricity delivered	Output	MWh	6,280,310	74,584,700	56,730	8,839,980
Total energy	Output	MWh	34,557,900	116,415,000	2339,000	26,752,600
Network length	Output	Miles	4064	16,292	1087	3253
Minimum temperature	Weather	° F	−10.35	19.90	−59.80	16.51
Wind speed	Weather	Knots	6.84	9.60	4.63	1.01
Precipitation	Weather	Inches	0.07	0.16	0.01	0.03
Growth in demand	Other	%	0.03	244.11	−74.96	17.72

[33], benchmarking of electricity transmission utilities is a challenging task due to the small number of transmission utilities that usually operate in the jurisdiction of a particular regulator. This likely explains why there are few empirical papers published on efficiency analysis of electricity transmission firms. Exceptions are [18–20,34]. However, none of these articles deal with unobserved heterogeneity or technological differences.[9]

The database used in this paper is as described in [3] and contains 405 observations on 59 US electricity transmission firms for the period 2001–2009. Following the literature, we specify a standard cost function with four outputs where our cost variable is *Totex* (which includes operation and maintenance expenses, annual depreciation on capital assets, and annual return on the balance of capital, and measured in 2000 dollars). The four outputs are: *Peak Load* (PL), which is the maximum peak load of the year during 60 min; *Electricity Delivered* (DE), which is the total annual energy delivered by the system; *Total Energy* (TE), which stands for the total energy of the system, including total net own generation, total purchases from others, net exchanges in the system (received–delivered), net transmission for others and transmission by others; and *Network length* (NL), which is a measure of the geographic spread of each company and is obtained as the sum of all transmission lines in miles regardless of the number of power cables on each power line. The four outputs considered (explanatory variables) and the cost variable (dependent variable) will be used later on in the DEA stage.

To analyse robustness, we extend the standard model by adding four time-invariant environmental variables to split the sample of transmission utilities. Three of these are weather variables: *Temperature* (TMIN), which represents the annual minimum temperature in Fahrenheit degrees; *Wind* speed (WIND), which is the average of the daily mean wind speeds in knots; and *Precipitation* (PRCP), which is the average of daily precipitation in inches. The last environmental variable is the *Growth in Demand* (GDEM) for each firm over time. The descriptive statistics of the full set of variables are shown in Table 3.

4.2. Empirical results

As above mentioned, we should initially use a simple specification of the cost function to split the sample in order to facilitate the replication of the procedure and to avoid convergence problems when more comprehensive models are estimated. In this sense, we use a Cobb–Douglas (or logarithm) specification of the cost function due to its widespread use and acceptance in previous empirical studies. Convergence problems prevented us from estimating the LCM for more than two classes with the linear specification that we used in our simulation exercise. However, these problems did not appear using the Cobb–Douglas functional form. As we do not know the true number of underlying technologies, this is an interesting advantage of the logarithm specification of the model. The coefficients for the Cobb–Douglas specification are shown in the Appendix.

Table 4
Efficiencies with the LCM–DEA procedure.

Number of classes	Average efficiency	Δ	% Obs. improv.
1	64.84	–	–
2	77.03	12.20	100.00
3	79.55	2.51	93.09
4	80.36	0.81	97.28
5	84.31	3.95	76.30
6	82.64	−1.66	56.79
7	86.71	4.06	69.14
8	87.51	0.80	61.23
9	87.41	−0.10	62.72

In Table 4 we show the descriptive statistics of the efficiency scores obtained using DEA as the number of classes is increased, and the number of observations (as a percentage) that improve their efficiency scores as we move from one class to two classes and so on. As expected, the average efficiency score for the so-called non-separation model, which can be considered as a model with one class, is 64.84%, much lower than the average efficiency obtained from the model with two classes, 77.03%. The most comprehensive model that is estimated is a LCM with 9 classes. Although the average efficiency score for this model goes up to 87.4%, the largest change in efficiencies occurs when we move from one class to two classes. We can also see that most observations have higher efficiency scores when more comprehensive models are estimated. This is compulsory for the 100% of the observations when we move from a model with one class to a model with two classes as the number of peers necessarily decreases in this case. This does not necessarily happen when we move from 2 to 3 classes (and so on) because there is a reallocation of the observations into the different classes and some observations might have "new" peers.[10]

The choice of the number of classes is a key issue in any clustering method. The AIC and BIC model selection criteria and some of their variants are commonly used to choose the appropriate number of classes in the LCM literature. The general form of most information criteria can be written as follows:

$$-2\ln LF + \text{Penalty} \tag{8}$$

where the first term is twice the negative logarithm of the maximum likelihood which decreases when the number of classes (complexity) increases. The penalty term penalizes too complex models, and increases with the number of parameters of the model.

[9] On the contrary there is an extensive literature in electricity distribution (see for instance, [35], for a European survey) and there are many articles that address the issue of heterogeneity including environmental factors in this sector (see for instance, [36–38], or [39]).

[10] See footnote #5.

Fig. 2. Choice of the number of classes.

Thus, these criteria involve minimizing an index that balances the lack of fit (too few classes) and overfitting (too many classes). Models with lower values of (8) are generally preferred.

Several information criteria are shown in Fig. 2 to illustrate robustness. The figure includes the traditional AIC and BIC criteria and some of their variants, the modified AIC criterion (AIC3), the corrected AIC (AICc), the so-called AICu, and the consistent AIC (CAIC) that can be considered either an AIC or BIC variant. For more details about these criteria and the associated penalty functions, see [40]. All of them show a remarkable improvement in fitness-parsimony when we move from just one class to a model with two classes. While the traditional AIC criterion and some of their variants (AIC3, AICc, and AICu) show little improvements when more classes are added, the BIC and CAIC clearly deteriorate their performance with more than two classes.[11] Generally speaking, the abovementioned tests allow us to conclude that a reasonable and practical trade-off between good description of the data and complexity is provided by a model with two classes. We therefore choose this model as our preferred model.

We also should expect notable differences in the composition of the two groups of firms if, instead of the LCM-based procedure, we use other procedures to split the sample. To give a sense of what difference this makes to the sample selection, we can pay attention to the percentage of observations for which this approach gives higher efficiencies than other methods. LCM provides not only the largest average efficiency score in the second stage, but also the majority of the observations obtain a higher value under this approach than from the others. More than 70% of observations are in an equal or better situation under LCM compared to the median of the network or cost and any cluster application. Thus, as discussed in the simulation section, the LCM approach provides the more favourable framework to benchmark firms. In addition, cluster-based separation procedures provide much more uneven sample partitions than the LCM approach. While in LCM there are 129 observations in a group and 276 in the other, under the two cluster applications the division is as follows: 49/356 when network is used as separating variable and 72/333 when also cost is included. This indicates the potential value to regulators of our LCM approach in reducing the need to rely on the small samples that can arise while using arbitrary approaches to sub-sample creation.

[11] The same happens if we use a criterion (not shown in Fig. 2) that penalizes poorly separated classes in LCMs with two or more classes, such as the so-called Complete Likelihood Classification (CLC) and the Integrated Classification Likelihood-BIC (ICL-BIC).

Table 5
Efficiencies obtained with different clustering methods.

Procedure	Mean	Std. Dev.	Max.	Min.
DEA (No separation)	64.84	21.87	100.00	9.15
LCM–DEA	77.03	19.22	100.00	9.39
Cluster (N)-DEA	65.52	22.45	100.00	9.15
Cluster (N, C)-DEA	67.05	21.23	100.00	11.77
Median of network-DEA	69.54	21.31	100.00	10.42
Median of cost-DEA	74.40	20.12	100.00	31.70

As a result of the above allocation, the alternative sample separation procedures provide different efficiency scores for each utility. The estimated efficiency levels are shown in Table 5. In accordance with the simulation results, the lowest levels are obtained not only when there is no separation of firms but also when we use cluster procedures using either network size or firms' cost as separating variables.[12] On the other hand, the largest efficiency scores are obtained when the LCM is used as a statistical tool to account for unobserved differences among firms. It is worth noting that most clustering procedures produce rather low efficiency scores for some observations. It should be noted, however, that this result has to do with application of DEA in the second stage. If we instead use a SFA approach, larger efficiencies would be obtained. This always happens because part of the measured inefficiency using DEA is now captured by the noise term of the model.

Using the median of cost as a sample-separating variable not only produces larger efficiency scores, but also a minimum efficiency (about 32%) that is much larger than in other clustering methods (including the LCM). This result is caused by the fact that we have used the same variable to both split the sample and measure firms' inefficiency. If we use the median of cost to split the sample we are falsely minimizing the differences in costs within each group. For instance, some very inefficient small firms (with relatively high costs) might be allocated with large firms, and some very efficient large firms (with relatively low costs) might be allocated with small firms. The consequence of these movements is both a balance of the average efficiencies of both classes, and an increase of the minimum efficiency level as the small (large) firms allocated with large (small) firms will become more efficient because the lack of peers with similar output levels. Generally speaking, the above discussion highlights the fact that we should not split the sample using a variable that is also being used to measure firms' inefficiency.

To give some intuition about the heterogeneity between classes that has been disentangled using the LCM procedure, we show in Table 6 the descriptive statistics of each one of the two groups that were found.[13] It can be seen that the average value of the cost and all the outputs is higher in class 1 than in class 2 so the largest companies are mainly located in the first class. However the standard deviations in class 1 for these variables are in general larger than in 2, indicating that there are more differences of size between firms in this class. Maybe this dissimilarity in the scale is because these firms operate in similar environments as it can be inferred from the smaller standard deviations of their environmental variables. The main difference on the average of these variables is observed for the temperature, indicating that in general firms of class 1 are located in colder regions, and the growth of the demand, which is mainly positive for firms in class

[12] The sample partition is the same when we take into account all the outputs and cost, or network size and cost together.

[13] To confirm the point made in footnote #1, we have checked the number of observations that freely change between classes in our model for the sample period. It can be observed that about 90% of our firms' observations remain in the same cluster from one year to another and hence no erratic changes are observed between classes over time.

Table 6
Descriptive statistics of the classes found with LCM.

Variable	Mean	Max.	Min.	Std. Dev.
	CLASS 1 (129 observations)			
TOTEX	224,218,000	667,127,000	20,713,600	166,005,000
PL	7730	23,111	380	6708
DE	6,376,580	39,484,700	82,304	7948,270
TE	40,417,200	115,685,000	2339,000	30,081,500
NL	5078	16,292	1088	4733
TMIN	−13.04	19.00	−59.80	14.05
WIND	6.81	9.22	4.63	1.05
PRCP	0.07	0.15	0.02	0.03
GDEM	0.80	62.59	−40.35	11.68
	CLASS 2 (276 observations)			
TOTEX	107,391,000	309,969,000	25,559,800	63,868,300
PL	5446	22,054	427	4730
DE	6235,310	74,584,700	56,730	9,240,720
TE	31,819,300	116,415,000	2,886,900	24,629,500
NL	3591	10,451	1087	2099
TMIN	−9.09	19.90	−59.80	17.43
WIND	6.85	9.60	4.76	1.00
PRCP	0.07	0.16	0.01	0.03
GDEM	−0.33	244.11	−74.96	19.94

Table 7
Clustering methods including environmental variables.

Procedure	Mean	Std. Dev.	Max.	Min.
LCM (W, D) - DEA	77.03	20.40	100.00	9.39
Cluster (W, D) - DEA	69.98	22.83	100.00	9.24
Cluster (W, D, N) - DEA	65.52	22.45	100.00	9.15
Cluster (W, D, N, C) - DEA	67.05	21.23	100.00	11.77

Table 8
Cobb–Douglas vs. Translog using LCM–DEA.

	Number of obs.		Av. eff.	
	CD	Translog	CD	Translog
	Without including environmental variables			
Class 1	129	102	66.83	63.61
Class 2	276	303	81.80	81.35
Both	405	405	77.03	76.88
	Including environmental variables			
Class 1	174	117	69.09	62.78
Class 2	231	288	83.01	82.25
Both	405	405	77.03	76.63

1 and negative for firms in class 2. These reasonable differences illustrate the nature of the heterogeneity that is controlled in our model. Clearly we could have arbitrarily allocated firms to two subsamples using a temperature threshold, however the LCM has allowed us to identify the number of subsamples to be analysed separately and then allocated firms between them in a statistically robust way.

4.3. Robustness analyses

We next introduce some additional tables in which we show the results obtained from alternative approaches or specifications that help us to analyse the robustness of the proposed clustering procedure based on LCM.

We first show in Table 7 the results we get when we introduce three weather variables and demand growth as sample-separating variables in the first stage of our procedure. Table 7 shows that both LCMs give us larger efficiency scores than extended k-means procedures that include environmental variables (alone or with information about the cost function). Based on our simulation results, we could then conclude that LCM also outperforms other sample separating methods when information about firms' environmental conditions is available. On the other hand, the estimated coefficients of the sample-separating variables (see Appendix) are statistically significant, which implies that they have helped to better split the sample. Despite this, our sample partition does vaguely change when we try to control for environmental variables as the percentage of coincidence in allocating observations is quite high (88%). This means that the between-class differences in estimated parameters are already capturing heterogeneity in firms' operating environment. In other words, a simple latent class model is able to control for those differences without explicitly including environmental variables that regulators might find it very difficult or expensive to collect. To examine better this issue, we have carried out an auxiliary regression (not shown) where an environmental variable composite interacts with the rest of explanatory variables of the frontier function. As we cannot reject that these coefficients are statistically significant, we can conclude that the parameter differences identified in a LCM model are, at least partially, capturing differences in environmental conditions.

Regarding the specification of the functional form, Table 8 provides a brief comparison of both Cobb–Douglas and Translog results. Again, the parameter estimates are shown in the Appendix. The correlation of Cobb–Douglas and Translog efficiency scores is very high (about 93%) and the overlap between classes is also remarkable (almost 84%). Furthermore, the model selection analysis indicates that the best trade-off between fitness and complexity for the Translog specification is provided once more by a model with two classes. A Cobb–Douglas specification is still preferred on the grounds of simplicity and because, in our case, the way in which the technology is modelled is not very relevant.

Although this paper does not attempt to contribute to the current debate about the suitability of parametric and nonparametric approaches for purposes of benchmark regulation, we now try to compare the relative performance of our procedure that combines LCM and DEA and two fully parametric procedures based on LCM, using both Cobb–Douglas and Translog specifications for the cost function. Besides the LCM–DEA approach, Table 9 provides results for the LCM–ALS model in which the traditional stochastic frontier model developed by Aigner et al. [41] (ALS) is applied in the second stage as done in Agrell et al. [2]. The third model, LCSFM, is a one-stage Latent Class Stochastic Frontier Model introduced by Greene et al. [6] that adds an inefficiency term to the LCM. Several interesting remarks are in order. First, when ALS is applied after the partition of the sample, the average efficiency for the whole sample is 100% because the estimated value of σ_u is equal to zero. This awkward result is often known as the 'wrong skewness problem' in the SFA literature, and might occur even when the model is correctly specified [42]. DEA likely became a very popular tool for benchmarking electric utilities because it allows regulators to address this issue. Second, LCSFM provides very similar partitions of the sample than the proposed procedure based on a non-frontier specification of the random term. For instance, the percentage of coincidences is about 98% when a Cobb–Douglas specification is used. This similarity is caused by the presence again of the 'wrong' skewness problem as the inefficiency term or σ_u (ignored in the proposed procedure) tends to vanish when a LCSFM is estimated. Therefore, it seems that a LCM model without a frontier specification and DEA is the best option to obtain proper efficiency levels in our application.

Finally, although the paper is focused on clustering methods, we also try to compare the relative performance of our proposed procedure and two non-clustering methods broadly used in the literature to take into account unobserved heterogeneity: the TFE and TRE models introduced by Greene [5,6]). As shown in Fig. 3, most of our earlier model selection criteria indicate that our empirical strategy based on estimating a LCM model provides a better fit than any of the stochastic frontier models introduced by Greene. This happens whether we use a Cobb–Douglas or Translog

Table 9
1 stage vs. 2 stages LCM clustering methods.

Specif.	Procedure	Class 1			Class 2		
		Number of obs.	Av. eff.	σ_u	Number of obs.	Av. eff.	σ_u
CD	LCM–DEA	129	66.83	–	276	81.80	–
	LCM–ALS	129	100.00	0.00	276	100.00	0.00
	LCSFM	138	87.28	0.32	267	100.00	0.00
Translog	LCM–DEA	102	63.61	–	303	81.35	–
	LCM–ALS	102	100.00	0.00	303	100.00	0.00
	LCSFM	67	77.21	0.52	338	100.00	0.00

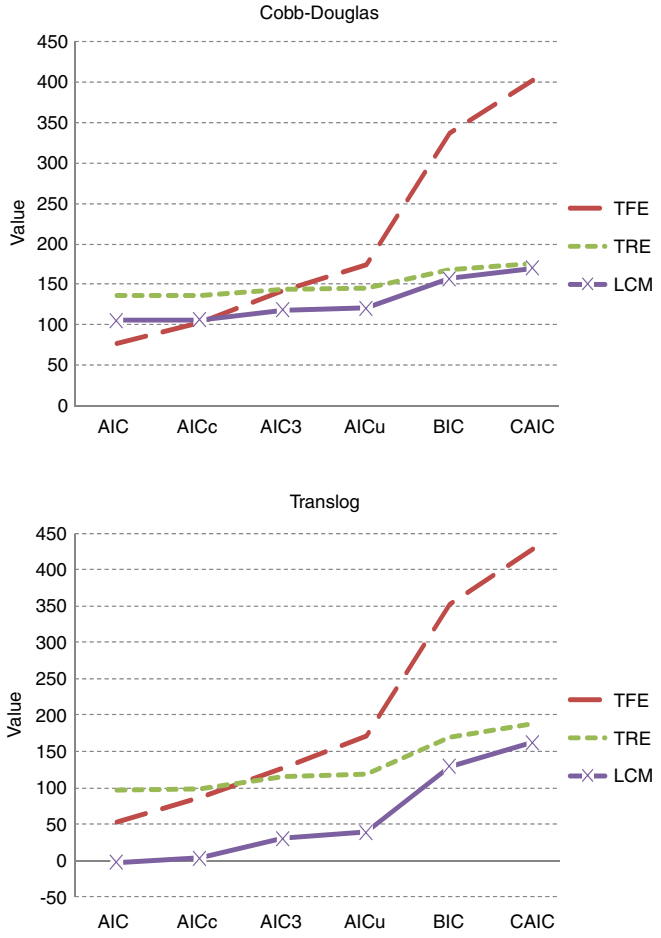

Fig. 3. Clustering vs. non-clustering model selection.

specification for the cost frontier. Our results seem to indicate that the underlying heterogeneity is better captured by a finite number of technologies rather than assuming that there are as many technologies as firms, but with the same marginal costs, economies of scale and other technological characteristics.

5. Conclusions

Electricity networks are often regulated by implementing incentive-based regulation schemes based on a comparison of utilities' performance with best-practice references. A key issue that is sometimes not taken into account is the heterogeneity or unobserved differences among firms associated with different technologies or environmental conditions. As in Agrell et al. [2], in this paper we propose using a latent class approach as a statistical clustering method to split the sample into groups of more comparable firms before carrying out a traditional efficiency analysis using DEA, the most common frontier analysis technique used by regulators in utility benchmarking.

We advocate this approach for several reasons. First, latent class models are specifically designed to cluster firms by searching for differences in production or cost parameters, which is exactly what regulators are looking for. Second, our approach can be viewed as a "supervised" method for clustering data as it takes into account the same relationship that is analysed later, often using nonparametric frontier techniques. And third, our approach is not more "technical" than other clustering methods as it can be implemented using standard software. The use of the same variables in both the latent class stage and the DEA stage and the use of simple model specifications contribute to simplifying the proposed procedure. We have demonstrated through a simulation exercise that the latent class approach better allocates observations into different classes than alternative clustering procedures and better predicts the underlying efficiency of each observation. The discriminatory capacity and the assignment success of the proposed clustering method increase when large differences between technologies or output distributions arise. This, in turn, yields a convergence of estimated efficiency levels to the true underlying levels. Moreover, the better the partition is, the larger the average efficiency scores are, whether we carry out either parametric or nonparametric efficiency analyses in the second stage of our procedure. From a regulation point of view, this outcome indicates that, given a number of classes, regulators could use the average efficiency level to compare the relative performance of several clustering methods in a real case in which they do not have information about the 'underlying partition' of the sample. In this sense, our simulation exercise justifies the use of a method that provides conservative efficiency estimates in benchmarking when, and only when, clustering methods are used.

Finally, we illustrate the proposed method with an application to a sample of US electricity transmission firms for the period 2001–2009. Several model selection tests allow us to conclude that a reasonable and practical trade-off between good description of the data and complexity is provided by a latent class model with two classes. In this sense, we also find that the largest change in efficiency scores occurs when we move from a one-class model (without any partition of the sample) to a model with only two classes. In line with our earlier simulation results, the largest efficiency scores are obtained when the LCM is used as a statistical tool to account for unobserved differences among firms.

We have also found that a simple latent class model is able to control for heterogeneity in firms' operating environment without explicitly including environmental variables that regulators might find it very difficult or expensive to collect. Our results seem to indicate that the underlying heterogeneity is better captured by a finite number of technologies (identified by a clustering method) than by using non-clustering methods that, in contrast, assume that there are as many technologies as firms, but with the same technological characteristics.

Acknowledgement

This research is partially funded by the Government of the Principality of Asturias through its predoctoral grants program. The authors also acknowledge the financial support of the Energy

Table A.1
Parameter estimates of the LCM using the US electricity transmission data.

Variable	Cobb–Douglas				Translog			
	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio
Class 1								
Constant	18.686	326.6	18.668	383.3	18.586	185.5	18.634	209.0
$\ln PL_{it}$	0.808	4.853	0.800	4.907	0.554	1.692	0.445	1.276
$\ln DE_{it}$	0.044	1.900	0.042	1.457	0.040	0.492	0.087	1.170
$\ln TE_{it}$	−0.261	−1.357	−0.237	−1.300	0.048	0.133	0.176	0.579
$\ln NL_{it}$	0.184	2.038	0.182	2.085	0.047	0.212	0.004	0.022
$1/2(\ln PL_{it})^2$					0.347	0.148	−0.754	−0.727
$1/2(\ln DE_{it})^2$					0.049	0.783	0.098	1.170
$1/2(\ln TE_{it})^2$					1.086	0.344	−0.894	−0.769
$1/2(\ln NL_{it})^2$					0.362	0.513	0.183	0.294
$\ln PL_{it} \cdot \ln DE_{it}$					0.521	2.002	0.327	1.295
$\ln PL_{it} \cdot \ln TE_{it}$					−0.678	−0.258	0.849	0.850
$\ln PL_{it} \cdot \ln NL_{it}$					0.015	0.024	−0.412	−0.731
$\ln DE_{it} \cdot \ln TE_{it}$					−0.567	−1.803	−0.485	−2.025
$\ln DE_{it} \cdot \ln NL_{it}$					0.055	0.413	0.100	0.795
$\ln TE_{it} \cdot \ln NL_{it}$					−0.142	−0.174	0.322	0.497
Sigma	0.380	22.982	0.381	22.078	0.356	15.218	0.332	15.719
Class 2								
Constant	18.385	1664.0	18.390	1649.0	18.227	1186.4	18.201	1164.2
$\ln PL_{it}$	0.144	3.109	0.166	3.881	0.273	6.622	0.423	8.491
$\ln DE_{it}$	0.054	5.258	0.060	5.823	0.048	5.113	0.043	4.069
$\ln TE_{it}$	0.415	7.817	0.401	7.785	0.295	5.794	0.106	1.754
$\ln NL_{it}$	0.136	6.192	0.123	5.133	0.164	9.013	0.182	8.158
$1/2(\ln PL_{it})^2$					0.440	2.403	1.519	5.569
$1/2(\ln DE_{it})^2$					0.066	6.968	0.024	2.071
$1/2(\ln TE_{it})^2$					0.017	0.060	1.624	3.870
$1/2(\ln NL_{it})^2$					0.487	9.009	0.4 71	7.533
$\ln PL_{it} \cdot \ln DE_{it}$					−0.142	−3.626	−0.126	−2.734
$\ln PL_{it} \cdot \ln TE_{it}$					−0.177	−0.796	−1.457	−4.398
$\ln PL_{it} \cdot \ln NL_{it}$					0.182	3.035	0.324	4.064
$\ln DE_{it} \cdot \ln TE_{it}$					0.090	1.789	0.093	1.649
$\ln DE_{it} \cdot \ln NL_{it}$					−0.041	−2.606	0.006	0.308
$\ln TE_{it} \cdot \ln NL_{it}$					−0.165	−2.127	−0.371	−3.318
Sigma	0.119	11.332	0.111	11.382	0.109	12.219	0.117	15.357
Class membership probabilities								
Constant			−0.088	−0.416			−0.881	−3.524
$TMIN_i$			−0.065	−3.001			−0.074	−3.255
$WIND_i$			−0.373	−2.153			0.974	3.215
$PRCP_i$			11.910	1.535			31.827	3.268
$GDEM_i$			0.092	1.744			−0.086	−1.582
Prior class prob.	0.444	0.556	0.479	0.521	0.351	0.649	0.292	0.708
Log LF	−39.666		−26.726		34.342		54.986	

Policy Research Group (University of Cambridge) and the Spanish Ministry of Science and Innovation through the project ECO2010-17240. Pollitt acknowledges the support of Ofgem in the collection of the US data.

Appendix

See Table A.1.

References

[1] Zhou P, Ang BW, Poh KL. A survey of Data Envelopment Analysis in energy and environmental studies. European J. Oper. Res. 2008;189:1–18.
[2] Agrell PJ, Farsi M, Filippini M, Koller M. 2013. Unobserved heterogeneous effects in the cost efficiency analysis of electricity distribution systems, CER-ETH Economics working papers series 13/171, CER-ETH - Center of Economic Research (CER-ETH) at ETH Zurich.
[3] Llorca M, Orea L, Pollitt MG. Efficiency and environmental factors in the US electricity transmission industry. Efficiency series papers 02/2013. Oviedo efficiency group. 2013.
[4] Haney AB, Pollitt MG. Efficiency analysis of energy networks: an international survey of regulators. Energy Policy 2009;37:5814–30.
[5] Greene W. Fixed and random effects in stochastic frontier models. J Prod Anal 2005;23:7–32.
[6] Greene W. Reconsidering heterogeneity in panel data estimator of the stochastic frontier model. J Econometrics 2005;126:269–303.
[7] Beard TR, Caudill SB, Gropper DM. Finite mixture estimation of multiproduct cost functions. The review of economics and statistics, vol. 73(4). MIT Press; 1991. p. 654–64.
[8] Gropper DM, Caudill SB, Beard TR. Estimating multiproduct cost functions over time using a mixture of normals. J Prod Anal 1999;11:201–18.
[9] Battese G, Rao DSP, O'Donnell CJ. A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. J Prod Anal 2004;21(1):91–103.
[10] O'Donnell CJ, Rao DSP, Battese G. Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. Empirical economics, vol. 34(2). Springer; 2008. p. 231–55.
[11] Verbeek M. A guide to modern econometrics. 3rd ed. Hoboken, NJ: John Wiley & Sons; 2008.
[12] Koehler AB, Murphree BS. A comparison of the Akaike and Schwarz criteria for selecting model order. Appl. Stat. 1988;37:187–95.
[13] Costello AB, Osborne JW. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Pract Asses Res Eval 2005;10(7):1–9.
[14] Johnson AL, Kuosmanen T. One-stage estimation of the effects of operational conditions and practices on productive performance: asymptotically normal and efficient, root-n consistent StoNEZD method. J Prod Anal 2011;36(2):219–30.
[15] Johnson AL, Kuosmanen T. One-stage and two-stage DEA estimation of the effects of contextual variables. European J. Oper. Res. 2012;220(2):559–70.
[16] Farrell MJ. The measurement of productive efficiency. J Roy Statist Soc Ser A 1957;120:253–81.
[17] Coelli TJ, Rao DSP, O'Donnell CJ, Battese GE. An introduction to efficiency and productivity analysis. 2nd ed. New York: Springer; 2005.
[18] Huettner DA, Landon JH. Electric utilities: scale economies and diseconomies. Southern Economic J 1978;44(4):883–912.
[19] Pollitt MG. Ownership and performance in electric utilities. Oxford: Oxford University Press; 1995.

[20] Dismukes DE, Cope III RF, Mesyanzhinov D. Capacity and economies of scale in electric power transmission. Utilities Policy 1998;7(3):155–62.
[21] Simar L, Wilson PW. Estimation and inference in two-stage, semi-parametric models of production processes. J Econometrics 2007;136(1):31–64.
[22] Po R-W, Guh Y-Y, Yang M-S. A new clustering approach using data envelopment analysis. European J. Oper. Res. 2009;199:276–84.
[23] Kruger JJ. Comment on ''A new clustering approach using data envelopment analysis''. European J. Oper. Res. 2010;206:269–70.
[24] Amin GR, Emrouznejad A, Rezaei S. Some clarifications on the DEA clustering approach. European J. Oper. Res. 2011;215:498–501.
[25] Moazami Goudarzi MR, Jaber Ansari MR. Clustering Decision Making Units (DMUs) using Full Dimensional Efficient Facets (FDEFs) of PPS with BCC technology. Appl Math Sci 2012;6(29):1431–52.
[26] Bogetoft P, Otto L. Benchmarking with DEA, SFA, and R. New York: Springer; 2011.
[27] Kuosmanen T, Saastamoinen A, Sipiläinen T. What is the best practice for benchmark regulation of electricity distribution? Comparison of DEA, SFA and StoNED Methods. Energy Policy 2013;61:740–50.
[28] Wang H-J, Schmidt P. One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. J Prod Anal 2002;18: 129–44.
[29] Kim M, Schmidt P. Valid tests of whether technical inefficiency depends on firm characteristics. J Econometrics 2008;144(2):409–27.
[30] Ruggiero J. Non-discretionary inputs in data envelopment analysis. European J. Oper. Res. 1998;111:461–9.
[31] Muñiz MA, Paradi J, Ruggiero J, Yang Z. Evaluating alternative DEA models used to control for non-discretionary inputs. Comput. Oper. Res. 2006;33(5): 1173–83.
[32] Agrell PJ, Bogetoft P. Development of benchmarking models for German electricity and gas distribution, Consultation report, Bundesnetzagentur, Bonn, Germany, 2007.
[33] Haney AB, Pollitt MG. International benchmarking of electricity transmission by regulators: theory and practice. Energy Policy 2013;62(November):267–81.
[34] von Geymueller P. Static versus dynamic DEA in electricity regulation: the case of US transmission system operators. CEJOR Cent. Eur. J. Oper. Res. 2009;17: 397–413.
[35] Jamasb T, Pollitt MG. International benchmarking and regulation: an application to European electricity distribution utilities. Energy policy, vol. 31(15). Elsevier; 2003. p. 1609–22.
[36] Yu W, Jamasb T, Pollitt MG. Does weather explain cost and quality performance? An analysis of UK electricity distribution companies. Energy Policy 2009;37(11):4177–88.
[37] Nillesen P, Pollitt MG. Using regulatory benchmarking techniques to set company performance targets: the case of US electricity. Compet Regul Netw Indus 2010;11(1):50–85.
[38] Jamasb T, Orea L, Pollitt MG. Estimating the marginal cost of quality improvements: the case of the UK electricity distribution companies. Energy Econ. 2012;34(5):1498–506.
[39] Growitsch C, Jamasb T, Wetzel H. Efficiency effects of observed and unobserved heterogeneity: evidence from Norwegian electricity distribution networks. Energy Econ. 2012;34(2):542–8.
[40] Fonseca JRS, Cardoso MGMS. Mixture-model cluster analysis using information theoretical criteria. Intelligent Data Analysis 2007;11(2):155–73.
[41] Aigner D, Lovell K, Schmidt P. Formulation and estimation of stochastic frontier production function models. J Econometrics 1977;6:21–37.
[42] Simar L, Wilson PW. Inferences from cross-sectional, Stochastic Frontier Models. Econometric Rev 2010;29(1):62–98.